



King's Research Portal

DOI:

[10.1109/ICC.2016.7511448](https://doi.org/10.1109/ICC.2016.7511448)

Document Version

Peer reviewed version

[Link to publication record in King's Research Portal](#)

Citation for published version (APA):

Wan Ariffin, W. N. S. F., Zhang, X., & Nakhai, M. R. (2016). Combinatorial Multi-armed Bandit Algorithms for Real-time Energy Trading in Green C-RAN. In *IEEE International Conference on Communications (IEEE ICC 2016)* IEEE. <https://doi.org/10.1109/ICC.2016.7511448>

Citing this paper

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

General rights

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Combinatorial Multi-armed Bandit Algorithms for Real-time Energy Trading in Green C-RAN

Wan Nur Suryani Firuz Wan Ariffin, Xinruo Zhang and Mohammad Reza Nakhai
Centre for Telecommunications Research, King's College London, Strand, UK, WC2R 2LS
E-mail: [k1206546, xinruo.zhang, reza.nakhai]@kcl.ac.uk

Abstract—Without a proper observation of the energy demand of the receiving terminals, the retailer may be obliged to purchase additional energy from the real-time market and may take the risk of losing profit. This paper proposes two combinatorial multi-armed bandit (CMAB) strategies in green cloud radio access network (C-RAN) with simultaneous wireless information and power transfer under the assumption that no initial knowledge of forthcoming energy demand and renewable energy supply are known to the central processor. The aim of the proposed strategies is to find the set of optimal sizes of the energy packages to be purchased from the day-ahead market by observing the instantaneous energy demand and learning from the behaviour of cooperative energy trading, so that the total cost of the retailer can be minimized. Two novel iterative algorithms, namely, *ForCMAB energy trading* and *RevCMAB energy trading* are introduced to search for the optimal set of energy packages in ascending and descending order of package sizes, respectively. Simulation results indicate that CMAB approach in our proposed strategies offers the significant advantage in terms of reducing overall energy cost of the retailer, as compared to other schemes without learning-based optimization.

I. INTRODUCTION

Cloud radio access network (C-RAN) has illustrated its considerable advantages in inter-cell interference (ICI) mitigation, system throughput improvement and reducing the capital expenditure and operating expense (OPEX) of the network operator, i.e., the retailer. Recently, the integration of C-RAN and simultaneous wireless information and power transfer (SWIPT), where the signals transmitted from remote radio heads (RRHs) can be exploited by the battery limited energy receiving terminals (ETs) for self-sustainability, has attracted the attention of researchers [1] and [2]. However, energy cost has become a major OPEX due to dramatic rise of energy consumption induced by the high density of RRHs deployment [3]. Consequently, the retailer may be obliged to purchase additional energy from the grid in the real-time market to cover extra energy requirements and may take a risk of losing profit. Subsequently, equipping the RRHs with renewable energy harvesting devices that can generate local renewable energy from environmental sources for green communications has been considered as a promising technique to benefit both the environment and the retailer. With the implementation of two-way energy trading with the grid, the retailer can maximally benefit from utilising their locally generated renewable energy and selling the excessive energy back to the grid [4]. Provided that all the base stations (BSs) are equipped with renewable energy harvesters and implemented with two-way energy trading, [4] proposes a joint energy trading and full cooperation scheme in coordinated multipoint (CoMP) network, where the data of

all the users is available at the central processor (CP) and will be distributed to all BSs for cooperative transmission via backhaul links. However, the data circulation between the CP and the BSs requires huge backhaul signalling overhead when full coordination is enabled [5]. This scheme, nevertheless, takes no consideration of backhaul capacity restrictions, which may be infeasible for practical capacity-constrained backhaul links.

Motivated by the literature that sparse beamforming problem is commonly formulated as a ℓ_0 -norm optimization problem and handled with reweighted ℓ_1 -norm method introduced in [6], the authors in [1], [7]–[10] propose dynamic sparse beamforming designs subject to quality of service (QoS) constraints for capacity-limited backhaul links in C-RAN. The authors in [1] integrate the aforementioned works with SWIPT concept and study the resource allocation algorithm. However, none of them take into account the renewable energy sources that can be further extended to the joint management of the resource allocation and energy trading for green communication. The joint cooperative resource management and energy trading problem in downlink green C-RAN are tackled in our previous works [2] and [11]. The former proposes a real-time BS energy management strategy in a green CoMP network with SWIPT, where the CP jointly minimizes the energy consumption and energy trading, in order to reduce the overall energy cost of the retailer. The latter studies three different cooperative real-time energy trading strategies in C-RAN and applies a sparse beamforming technique to find the optimal trade-off between the degree of partial cooperation among the RRHs in serving the receiving terminals and the total energy cost of the retailer. However, in both papers, we assume that the retailer purchases a set of fixed sizes of energy packages for the RRHs from the day-ahead market without considering the actual amount of local renewable energy generation at RRHs as well as the changes in energy requirements of the information receiving terminals (ITs) and ETs.

In this paper, we further extend our previous works to a learning-based practical approach and assume that the CP has no initial knowledge of forthcoming power budget and energy consumption at the individual RRHs. With various sizes of energy packages which are offered in the day-ahead market, the new responsibility of the CP is to determine the set of optimal sizes of the energy packages to be purchased for the RRHs on the basis of actual energy supply and demand, to further minimize the total energy cost of the retailer. We model this problem as a combinatorial multi-armed bandit (CMAB) problem and introduce two algorithms, namely, *ForCMAB Energy Trading* and *RevCMAB Energy Trading*, to observe the

energy demand and search for the set of optimal sizes of the energy packages to be purchased from the day-ahead market.

The rest of this paper is organized as follows. Section II introduces the system model. In section III, a real-time cooperative energy trading strategy is formulated and then transformed into numerically tractable form. In section IV, two combinatorial multi-armed bandit algorithms are proposed. Numerical simulation results are analyzed in section V. Finally, section VI summarizes the paper. *Notations:* w , \mathbf{w} , \mathbf{W} , $(\cdot)^H$ and $\text{tr}(\cdot)$, respectively, represent a scalar w , a vector \mathbf{w} , a matrix \mathbf{W} , the complex conjugate transpose operators and the trace operators. $\mathbf{W} \succeq 0$ denotes that \mathbf{W} is a positive semidefinite matrix and $\mathbb{C}^{n \times m}$ indicates the sets of n -by- m dimensional complex matrices. $\|\cdot\|_p$ denotes the ℓ_p -norm of a vector and $\|\cdot\|_0$ indicates the number of non-zero entries in the vector. Note that, the normalized energy unit, i.e., J_s^{-1} , is adopted in this paper, therefore, the terms 'energy' and 'power' are mutually convertible.

II. SYSTEM MODEL

We consider a downlink C-RAN with SWIPT from N RRHs, equipped with M antennas, towards K_i active single-antenna ITs and K_e active single-antenna ETs, respectively, over a shared bandwidth. The CP is the central processing unit that coordinates all the cooperative strategies for the RRHs based on perfect knowledge of channel state information and distributes data of all the ITs to the corresponding RRHs via backhaul links. Let $\mathcal{L}_b = \{1, \dots, N\}$, $\mathcal{L}_e = \{1, \dots, K_e\}$, $\mathcal{L}_e^{\text{idle}} = \{1, \dots, K_e^{\text{idle}}\}$ and $\mathcal{L}_i = \{1, \dots, K_i\}$ indicate, respectively, the set of indexes of the RRHs, the active ETs, the idle ETs and the active ITs. We further assume that the energy transmission between the CP and RRHs is accomplished via dedicated power lines.

A. Energy Management model

We assume that at least one renewable energy harvesting device, e.g., wind turbine and/or solar panel, is installed at each RRH in order to generate local renewable energy from environmental sources. Furthermore, no RRH is equipped with frequently rechargeable storage devices and the RRHs are obliged to transmit the excessive power back to the CP for sale. In practice, the renewable energy generation is unequal due to efficiency of different types of renewable energy harvesting devices and various RRHs locations. Let E_n , B_n^{ahead} , B_n^{real} , S_n be defined, respectively, as the amount of renewable energy generated at the n -th RRH, the amount of energy that has already been purchased from the grid in the day-ahead market, the amount of energy that is necessary to be purchased from the real-time market and the amount of excessive energy sold back to the grid via the retailer. Furthermore, let P_n^{Tx} and P_n^{circuit} indicate the total transmit power at the n -th RRH and the hardware circuit power consumption at the n -th RRH, respectively. Then, the total energy consumption at the n -th RRH, i.e., P_n^{total} , is upper-bounded by the total available energy at the n -th RRH, i.e.,

$$P_n^{\text{total}} = P_n^{\text{Tx}} + P_n^{\text{circuit}} \leq E_n + B_n^{\text{ahead}} + B_n^{\text{real}} - S_n. \quad (1)$$

In practice, the price of generating per unit renewable energy, denoted by π^{renew} , is much cheaper than the price of purchasing per unit energy from the day-ahead market, represented

by π^{ahead} . From the supply and demand perspective, we assume that the retailer purchases additional energy from the real-time market at a higher price π^{real} and sells the unconsumed energy back to the grid at a reduced price π^{sell} , i.e., $\pi^{\text{real}} \geq \pi^{\text{ahead}} \geq \pi^{\text{sell}} \geq \pi^{\text{renew}}$.

B. Downlink Transmission Model

We denote the overall beamforming vector from all the RRHs towards the i -th IT, $i \in \mathcal{L}_i$, as $\mathbf{w}_i = [\mathbf{w}_{1i}^H, \dots, \mathbf{w}_{Ni}^H]^H \in \mathbb{C}^{MN \times 1}$, where $\mathbf{w}_{ni} \in \mathbb{C}^{M \times 1}$ is the beamformer from the n -th RRH towards the i -th IT. Let $\mathbf{v}_e = [\mathbf{v}_{1e}^H, \dots, \mathbf{v}_{Ne}^H]^H \in \mathbb{C}^{MN \times 1}$ represent the overall beamforming vector from all the RRHs to the e -th active ET and $\mathbf{h}_{ni} \in \mathbb{C}^{M \times 1}$ indicate the channel vector between the n -th RRH and the i -th IT. We denote the overall channel vector between all the RRHs and the i -th IT by $\mathbf{h}_i = [\mathbf{h}_{1i}^H, \dots, \mathbf{h}_{Ni}^H]^H \in \mathbb{C}^{MN \times 1}$. The received signals at the i -th IT, $i \in \mathcal{L}_i$, is then given by

$$y_i = \mathbf{h}_i^H \mathbf{w}_i s_i^{\text{IT}} + \sum_{\substack{j \neq i \\ j \in \mathcal{L}_i}} \mathbf{h}_i^H \mathbf{w}_j s_j^{\text{IT}} + \sum_{e \in \mathcal{L}_e} \mathbf{h}_i^H \mathbf{v}_e s_e^{\text{ET}} + n_i, \quad (2)$$

where the terms at the right hand side of (2), respectively, represent the information-carrying signal intended to the i -th IT, the inter-user interference caused by all other non-desired information beams, the interference caused by the energy-carrying signals intended to all active ETs and the additive white Gaussian noise with variance of σ_i^2 at the i -th IT. Since no information is carried by the energy-carrying signals, they can be any arbitrary random signals generated at RRHs. We assume that the transmitted symbols, i.e., s_i^{IT} , s_j^{IT} and s_e^{ET} , are independent and identically distributed and their transmission energy is normalized to one. Without loss of generality, we also let the noise variances σ_i^2 be identical at all receiving terminals. Then, the signal-to-interference-plus-noise ratio (SINR) at the i -th IT, $i \in \mathcal{L}_i$, is formulated as

$$\text{SINR}_i^{\text{IT}} = \frac{|\mathbf{h}_i^H \mathbf{w}_i|^2}{\sum_{j \in \mathcal{L}_i, j \neq i} |\mathbf{h}_i^H \mathbf{w}_j|^2 + \sum_{e \in \mathcal{L}_e} |\mathbf{h}_i^H \mathbf{v}_e|^2 + \sigma_i^2}. \quad (3)$$

The backhaul capacity consumption for the n -th RRH can be expressed as $C_n^{\text{backhaul}} = \sum_{i \in \mathcal{L}_i} \|\mathbf{w}_{ni}\|_0^2 R_i$, $\forall n \in \mathcal{L}_b$,

where $R_i = \log_2(1 + \gamma_i)$ is the achievable data rate (bit/s/Hz) by the i -th IT and $\|\mathbf{w}_{ni}\|_0^2$ is an indicator function that illustrates the scheduling choices of the individual ITs, i.e., the i -th IT here as

$$\|\mathbf{w}_{ni}\|_0^2 = \begin{cases} 0, & \text{if } \|\mathbf{w}_{ni}\|_2^2 = 0, \\ 1, & \text{if } \|\mathbf{w}_{ni}\|_2^2 \neq 0. \end{cases} \quad (4)$$

Here, $\|\mathbf{w}_{ni}\|_2^2 = 0$ indicates that the i -th IT is not served by/scheduled to the n -th RRH and, hence, the backhaul link between the CP and the n -th RRH is not used for joint data transmission to the i -th IT. The total energy harvested by the e -th active ET, $e \in \mathcal{L}_e$ can be defined as

$$\mathcal{G}_e^{\text{ET}} = \eta \left(|\mathbf{g}_e^H \mathbf{v}_e|^2 + \sum_{j \in \mathcal{L}_e, j \neq e} |\mathbf{g}_e^H \mathbf{v}_j|^2 + \sum_{i \in \mathcal{L}_i} |\mathbf{g}_e^H \mathbf{w}_i|^2 \right), \quad (5)$$

where $0 \leq \eta \leq 1$ indicates the conversion efficiency from the harvested RF energy to the electrical energy, $\mathbf{g}_e =$

Algorithm 1 summarizes the steps of sparse beamforming design for a given set of energy package purchased from day-ahead market, where the problem in (8) is repeatedly solved using the iteratively updated weight factor ξ_{ni} and κ_{ne} . The cooperative links between the RRHs and the active receiving terminals are iteratively removed on the basis of the power budgets and backhaul link capacity restrictions at the individual RRHs. Consequently, the RRHs with low transmit power in the

k -th iteration result in high weight factors, which will further force the transmit power to be reduced in the $(k+1)$ -th iteration until the solution sparsity is attained.

Algorithm 1 Reweighted ℓ_1 -norm method.

- 1: **Initialize:** constant $\mu \rightarrow 0$, iteration count $k = 0$, weight factor $\xi_{ni}(0) = 1$, $\kappa_{ne}(0) = 1$, maximum number of iterations k_{max} .
 - 2: **while** ξ_{ni} and κ_{ne} are not converged or $t \neq k_{max}$ **do**
 - 3: Find the optimal $\mathbf{W}_i(k)$ and $\mathbf{V}_e(k)$ by solving (8);
 - 4: Update the weight factor $\xi_{ni}(k+1)$ as follows,

$$\xi_{ni}(k+1) = \frac{1}{\text{tr}(\mathbf{W}_i(k)\mathbf{D}_n) + \mu}, \quad \forall n \in \mathcal{L}_b, i \in \mathcal{L}_i;$$
 - 5: Update the weight factor $\kappa_{ne}(k+1)$ as follows,

$$\kappa_{ne}(k+1) = \frac{1}{\text{tr}(\mathbf{V}_e(k)\mathbf{D}_n) + \mu}, \quad \forall n \in \mathcal{L}_b, e \in \mathcal{L}_e;$$
 - 6: Increment the iteration number $k = k + 1$;
 - 7: **end while**
-

IV. COMBINATORIAL MULTI-ARMED BANDIT FOR REAL-TIME ENERGY TRADING IN C-RAN

The multi-armed bandit (MAB) problem models a slot machine attempts to maximize the accumulated reward by iteratively optimizing the decisions among a set of arms based on existing knowledge, known as exploitation, while simultaneously acquiring new knowledge by observing the associated reward, known as exploration [12].

In this paper, we employ an abstract idea of MAB for the CP to learn from the behaviour of the energy trading in C-RAN iteratively, where each arm corresponds to the size of an energy package per RRH to be purchased by the retailer from the day-ahead market prior to the actual time of energy demand. Therefore, the new responsibility of the CP is to find the set of optimal sizes of the energy packages to be purchased from the day-ahead market without the initial knowledge of forthcoming instantaneous energy demand, in order to minimize the total energy cost of the retailer. We assume that there is a total number of \mathcal{J} arms, where only N arms, $N \subset \mathcal{J}$, can be pulled simultaneously. At each trial, the CP chooses N sizes of energy packages for N RRHs, which is equivalent to pulling a set of N arms simultaneously. Then, the CP observes the individual reward for each arm and calculates the aggregated reward. This problem falls in the category of combinatorial MAB (CMAB) [13], because multiple arms are pulled simultaneously and the reward for each arm is observed individually. Let $\mathcal{A}_t^{[set]} = \{B_1^{[ahead]}(t), \dots, B_N^{[ahead]}(t)\}$ indicates N sizes of energy packages to be purchased by the retailer from the day-ahead market at the t -th trial. The problem is to decide which combination of arms should be pulled at each t -th trial in order to maximize the individual reward for the n -th RRH, i.e., $\mathcal{R}(B_n^{[ahead]}(t))$ and, thus, the accumulated reward, i.e., $\mathcal{R}(\mathcal{A}_t^{[set]})$ in a limited number of trials T , which can be calculated as

$$\mathcal{R}(B_n^{[ahead]}(t)) = B_n^{[total]}(0) - B_n^{[total]}(t), \quad \forall n \in \mathcal{L}_b, \quad (9)$$

$$\mathcal{R}(\mathcal{A}_t^{[set]}) = \sum_{n \in \mathcal{L}_b} \mathcal{R}(B_n^{[ahead]}(t)). \quad (10)$$

$B_n^{[total]}(0)$ and $B_n^{[total]}(t)$ in (9) are the total energy cost of the retailer consumed by the n -th RRH at the initial trial, i.e.,

$t = 0$ and at the t -th trial respectively, which are given by

$$B_n^{[total]}(t) = \pi^{[ahead]} B_n^{[ahead]}(t) + \pi^{[real]} B_n^{[real]}(t) + \pi^{[renew]} E_n(t) - \pi^{[sell]} S_n(t), \quad \forall n \in \mathcal{L}_b. \quad (11)$$

In the sequel, we introduce a *forward CMAB (ForCMAB) Energy Trading* algorithm to find the optimal combination of arms in an ascending order of packages sizes in a given number of trials T , where $T = t_{exp} + t_{max}$ is the total of exploration trials, t_{exp} and exploitation trials, t_{max} . The steps are summarized in Algorithm 2, where we assume a set of index of the sizes of energy packages offered by the grid during the day-ahead market $\mathcal{E}_t^{[total]} = \{\mathcal{E}^1, \dots, \mathcal{E}^{\mathcal{J}}\}$ with $\mathcal{E}^1 < \mathcal{E}^2 < \dots < \mathcal{E}^{\mathcal{J}}$ is an arithmetic progression (AP) with common difference of \mathcal{C} . Furthermore, we define package^[size] = $x\mathcal{C}$, where x is a constant number that has been chosen by the CP.

Algorithm 2 ForCMAB Energy trading

- 1: **Initialize:** $\mathcal{A}_0^{[set]} = \{0_1, \dots, 0_N\}$, exploration count $t = 0$, exploitation count $u = 0$, maximum number of trials T .
- 2: Run Algorithm 1;
 CP calculates $B_n^{[total]}(0)$ and defines the initial individual reward $\mathcal{R}(B_n^{[ahead]}(0)) = 0$, accumulated reward $\mathcal{R}(\mathcal{A}_0^{[set]}) = 0$ and total accumulated reward over T

$$\sum_{x=1}^T \mathcal{R}(\mathcal{A}_x^{[set]}) = 0, \quad \forall n \in \mathcal{L}_b$$
- 3: **while** $\mathcal{R}(\mathcal{A}_t^{[set]}) \geq \mathcal{R}(\mathcal{A}_{t-1}^{[set]})$ **do**
- 4: Increment the iteration number $t = t + 1$.
- 5: Run Algorithm 1
- 6: **Exploration** of new combinatorial arms, $\mathcal{A}_t^{[set]}$ by solving (8);
- 7: CP calculates $B_n^{[total]}(t)$, $\mathcal{R}(B_n^{[ahead]}(t))$ and $\mathcal{R}(\mathcal{A}_t^{[set]})$, $\forall n \in \mathcal{L}_b$
- 8: **If** the individual reward for the n -th RRH

$$\mathcal{R}(B_n^{[ahead]}(t)) \geq \mathcal{R}(B_n^{[ahead]}(t-1)) \text{ and } B_n^{[ahead]}(t) \neq \mathcal{E}^{\mathcal{J}}, \quad \forall n \in \mathcal{L}_b$$
- 9: **then** update $B_n^{[ahead]}(t+1) = B_n^{[ahead]}(t) + \text{package}^{[size]}$,
- 10: **else** update $B_n^{[ahead]}(t+1) = B_n^{[ahead]}(t)$, $\forall n \in \mathcal{L}_b$.
- 11: **end if**
- 12: Update $\mathcal{A}_{t+1}^{[set]} = \{B_1^{[ahead]}(t+1), \dots, B_N^{[ahead]}(t+1)\}$.
- 13: **end while**
- 14: **Set** $\mathcal{A}_\star^{[set]} = \mathcal{A}_{t+1}^{[set]} = \{B_1^{[ahead]}(t+1), \dots, B_N^{[ahead]}(t+1)\}$.
- 15: **Update** $t_{exp} = t$
- 16: **while** $u \neq T$ **do**
- 17: Increment the iteration number $u = t_{exp} + 1$.
- 18: **Exploitation** of the arms with the optimal reward $\mathcal{A}_u^{[set]} = \mathcal{A}_\star^{[set]}$, run Algorithm 1 and solve (8).
- 19: CP calculates $B_n^{[total]}(u)$, $\mathcal{R}(B_n^{[ahead]}(u))$ and $\mathcal{R}(\mathcal{A}_u^{[set]})$
- 20: **end while**
- 21: Calculate total accumulated reward over T ;

$$\sum_{x=1}^T \mathcal{R}(\mathcal{A}_x^{[set]}) = \sum_{t=1}^{t_{exp}} \mathcal{R}(\mathcal{A}_t^{[set]}) + \sum_{u=t_{exp}+1}^T \mathcal{R}(\mathcal{A}_u^{[set]})$$

Note that in Algorithm 2, the CP starts the operation with $\mathcal{A}_0^{[set]} = \{0_1, \dots, 0_N\}$ to check if the available local renewable energy is sufficient to satisfy the QoS of the ETs and ITs. If satisfied, the CP will decide not to buy any packages from

the day-ahead market. At each trial t , the CP decides to purchase N combination of sizes of the energy packages from the day-ahead market for N RRHs based on the individual reward obtained from the current trial t and the previous trial $(t - 1)$. The iterations of CMAB search are continued until the aggregated reward is maximized. In contrast to the strategy used in the former algorithm, we introduce a *reverse CMAB (RevCMAB) Energy Trading* algorithm, i.e., Algorithm 3 to find the optimal combination of the energy packages in descending order of packages sizes to be purchased from the day-ahead market, so that the maximum individual RRH reward can be achieved, and thus, the total energy cost of the retailer is minimized.

Algorithm 3 *RevCMAB Energy trading*

- 1: **Initialize:** a set of maximum sizes of energy packages offered by the day-ahead market $\mathcal{A}_0^{[\text{set}]} = \{\mathcal{E}_1^{\mathcal{J}}, \dots, \mathcal{E}_N^{\mathcal{J}}\}$.
 - 2: STEP 2 - STEP 7 in Algorithm 2
 - 3: **If** the individual reward for the n -th RRH $\mathcal{R}(B_n^{[\text{ahead}]}(t)) \geq \mathcal{R}(B_n^{[\text{ahead}]}(t-1))$ and $B_n^{[\text{ahead}]}(t) \neq 0$, $\forall n \in \mathcal{L}_b$
 - 4: **then** update $B_n^{[\text{ahead}]}(t+1) = B_n^{[\text{ahead}]}(t) - \text{package}^{[\text{size}]}$,
 - 5: **else** update $B_n^{[\text{ahead}]}(t+1) = B_n^{[\text{ahead}]}(t)$, $\forall n \in \mathcal{L}_b$.
 - 6: **end if**
 - 7: STEP 12 - STEP 21 in Algorithm 2
-

V. SIMULATION RESULTS

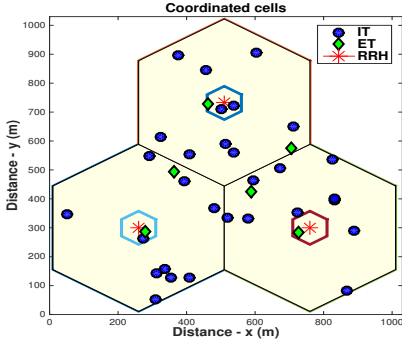


Fig. 1. A Multi-user Downlink SWIPT C-RAN Simulation Topology.

We consider a downlink C-RAN consists of 3 adjacent RRHs SWIPT towards 30 single-antenna ITs and 6 single-antenna ETs. Each RRH is equipped with 8 antennas and located 500m away from each other, as shown in Fig 1. The renewable energy generation at each RRH is assumed to be $E_1 = 1.5$ W, $E_2 = 0.2$ W and $E_3 = 0.05$ W, respectively, at the price of $\pi^{[\text{renew}]} = \text{£}0.02/\text{W}$. The retailer has purchased a set of energy packages $\mathcal{A}^{[\text{set}]} = \{B_1^{[\text{ahead}]}, B_2^{[\text{ahead}]}, B_3^{[\text{ahead}]}\}$ W from the day-ahead market at the price of $\pi^{[\text{ahead}]} = \text{£}0.07/\text{W}$. We run the Algorithm 2 and Algorithm 3 for $T = 24$ trials with $\mathcal{J} = 48$ and $\mathcal{E}_i^{[\text{total}]} = \{100, 200, \dots, 4800\}$ mW with common difference of $\mathcal{C} = 100$ mW. We further assume that the retailer can purchase additional energy from the real-time market at the price of $\pi^{[\text{real}]} = \text{£}0.15/\text{W}$ and sell excessive energy back to the grid at the price of $\pi^{[\text{sell}]} = \text{£}0.05/\text{W}$. A correlated channel

model $\mathbf{h}_{ni} = \mathbf{R}^{1/2} \mathbf{h}_w$ is adopted [14], where $\mathbf{h}_w \in \mathbb{C}^{M \times 1}$ are zero-mean circularly symmetric complex gaussian (ZMCSG) random variables with unit variance, $\mathbf{R} \in \mathbb{C}^{M \times M}$ is the spatial covariance matrix and its (m, n) -th element is given by $G_a L_p \sigma_F^2 e^{-0.5 \frac{(\sigma_s \ln 10)^2}{100}} e^{j \frac{2\pi \delta}{\lambda} [(n-m) \sin \theta]} e^{-2 \left[\frac{\pi \delta \sigma}{\lambda} (n-m) \cos \theta \right]^2}$, where $G_a = 15\text{dBi}$ is antenna gain, $L_p(\text{dB}) = 125.2 + 36.3 \log_{10}(d)$ is the path loss model over a distance of d km, σ_F^2 is the variance of the complex Gaussian fading coefficient, $\sigma_s = 8$ dB is the log-normal shadowing standard deviation, $\delta = \lambda/2$ is the antenna spacing, $\sigma = 2^\circ$ is the angular offset standard deviation and θ is the estimated angle of departure. The channel bandwidth, noise figure at receiving terminals and noise power spectral density are set to be 20 MHz, 5 dB and -174 dBm/Hz, respectively. Equal weight factor of serving area for ETs is assumed to be $\wp_{nm} = 0.2$. Besides, the parameters for optimization constraints are set to be $P_{\text{CP}}^{[\text{circuit}]} = 40$ dBm, $P_{\text{CP}}^{[\text{max}]} = 50$ dBm, $P_n^{[\text{circuit}]} = 30$ dBm, $P_n^{[\text{Tmax}]} = 46$ dBm, $C_n^{[\text{b-limit}]} = 1200$ bits/s/Hz, $P_e^{[\text{min}]} = -60$ dBm, $P_z^{[\text{idle}]} = -90$ dBm and $\eta = 0.5$, respectively. The simulation results are efficiently obtained and averaged over 100 independent channel realizations via CVX [15]. In order to demonstrate the advantages of our proposed strategies, the strategy in [2] and [11] that assume a set of fixed energy packages, i.e., $\mathcal{A}^{[\text{set}]} = \{B_1^{[\text{ahead}]} = B_2^{[\text{ahead}]} = B_3^{[\text{ahead}]} = 700 \text{ mW}\}$, are employed in this paper as comparison group and identical constraints are applied to all the strategies for fair comparison. Furthermore, we employ different package sizes for the proposed strategies to study the impact of system parameters. They are, respectively, *ForCMAB Energy Trading* algorithm with package^[size] = 200 mW, *ForCMAB Energy Trading* algorithm with package^[size] = 100 mW, and *RevCMAB Energy Trading* algorithm with package^[size] = 200 mW.

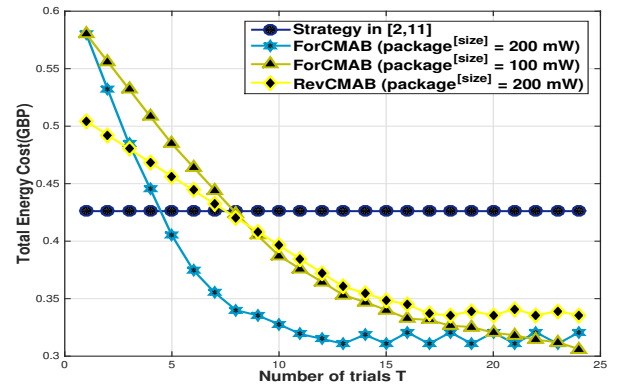


Fig. 2. Total energy cost versus number of trials at $\gamma = 20\text{dB}$.

Fig. 2 compares the total energy cost of the retailer versus number of trials for different strategies at $\gamma = 20$ dB. One can conclude that conducive to reducing the total energy cost, overwhelming performance gain can be achieved by both of the proposed CMAB strategies after a few number of trials as compared to the strategy proposed in [2] and [11] that assume a fixed set of energy packages $\mathcal{A}^{[\text{set}]}$ over the trials. The performance gap in the first number of trials is due to the exploration in the learning process of our proposed CMAB strategies.

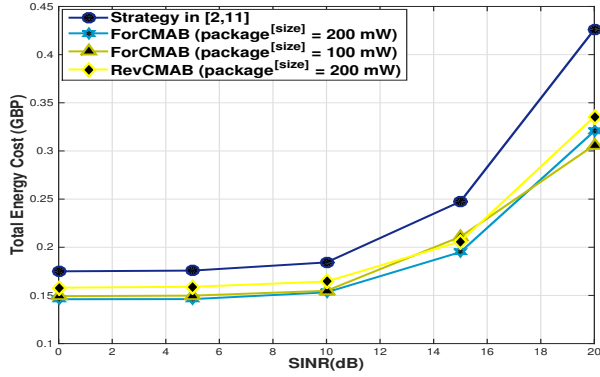


Fig. 3. Total energy cost versus various target SINR at $T = 24$.

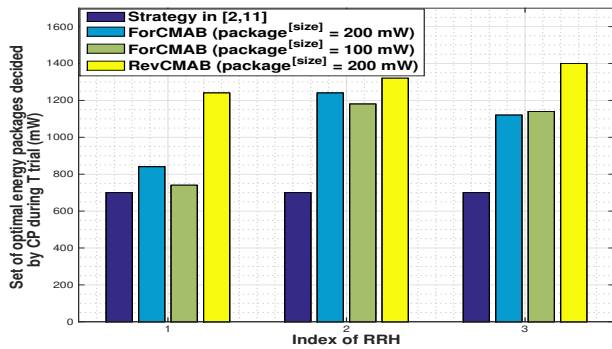


Fig. 4. Average set of optimal sizes of the energy packages to be purchased from the day-ahead market decided by the CP at $T = 24$ and $\gamma = 20$ dB.

As it can be observed from the figure, the learning speed of the *ForCMAB Energy Trading* algorithm is much higher than *RevCMAB Energy Trading* algorithm for package^[size]=200 mW. The comparison of the total energy cost of the retailer versus various SINR targets for different strategies at the 24-th trial is illustrated in Fig. 3. It can be observed from the figure that after 24 times of trials, both of our proposed *ForCMAB Energy Trading* and *RevCMAB Energy Trading* algorithms outperform the strategy proposed in [2] and [11] in terms of total energy cost of the retailer. Furthermore, the performance gap increases with the increasing SINR requirements. Fig. 4 presents in details a set of the optimal energy packages chosen by the CP to be purchased from the day-ahead market at the 24th trial. It is noticeable that instead of using a fixed set of energy packages, both of our proposed strategies purchase a set of optimal energy packages from day-ahead market on the basis of actual energy generation and the energy requirements at the individual RRHs. In addition, even though both of the proposed strategies have similar performance in terms of total energy cost of the retailer at $\gamma = 20$ dB, the proposed *RevCMAB Energy Trading* algorithm tends to purchase higher amount of energy packages from the day-ahead market for the individual RRHs.

VI. CONCLUSION

This paper proposes two CMAB algorithms, namely, *ForCMAB Energy Trading* and *RevCMAB Energy Trading*, to

observe the instantaneous energy demand and learn from the behaviour of cooperative energy trading in the green C-RAN with SWIPT. At each trial, the sparse beamforming technique is employed to find the optimal trade-off between the cooperative transmission and the total energy cost of the retailer. Assuming that the RRHs have no initial knowledge of forthcoming energy consumption and renewable energy production in real-time energy trading, we employed the CMAB learning process to search for the optimal set of energy packages to be purchased from the day-ahead market in a limited number of trials, to further reduce the total energy cost of the retailer. Our simulation results confirm that in terms of reducing the total energy cost of the retailer, both proposed strategies outperform two recently proposed strategies without CMAB approach in green C-RAN.

REFERENCES

- [1] D. W. K. Ng and R. Schober, "Resources Allocation for Coordinated Multipoint Networks with Wireless Information and Power Transfer", in *IEEE Global Communications Conference*, Dec. 2014.
- [2] Wan Ariffin, W.N.S.F., Zhang, X. and Nakhai, M.R., "Real-time Power Balancing in Green CoMP Network with Wireless Information and Energy Transfer", in *IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, pp. 1574 - 1578, Sept. 2015.
- [3] A. Fehske *et al.*, "The Global Footprint of Mobile Communications: The Ecological and Economic Perspective," in *IEEE Communications Magazine*, pp. 5562, August 2011.
- [4] J. Xu and R. Zhang, "Cooperative Energy Trading in CoMP Systems Powered by Smart Grids", in *IEEE Global Communications Conference*, pp. 2697-2702, Dec. 2014.
- [5] T. Le, S. Nasser, A. Zarrebini-esfahani, M. R. Nakhai, A. Mills, "Power-efficient downlink transmission in multicell networks with limited wireless backhaul, *IEEE Wireless Communications*, vol. 18, no. 5, pp. 82-88, Oct. 2011.
- [6] E. Candes, M. Wakin and S. Boyd, "Enhancing Sparsity by Reweighted ℓ_1 Minimization," in *Journal of Fourier Analysis and Applications*, vol.14, no.5, pp. 877-905, 2008.
- [7] B. Dai *et al.*, "Sparse Beamforming and User-Centric Clustering for Downlink Cloud Radio Access Network," in *IEEE Access:Recent Advances in Cloud RAN*, vol. 2, pp. 1326-1339, Oct. 2014.
- [8] J. Zhao *et al.*, "Coordinated Multipoint Transmission with Limited Backhaul Data Transfer," in *IEEE Trans. Wireless Communications*, vol. 12, pp. 2762-2775, Jun. 2013.
- [9] D.W.K.Ng, E.S.Lo, and R.Schober, "Energy-Efficient Resource Allocation in Multi-Cell OFDMA Systems with Limited Backhaul Capacity," in *IEEE Trans. Wireless Commun.*, vol.11, pp. 3618-3631, Oct.2012.
- [10] M. Hong *et al.*, "Joint Base Station Clustering and Beamformer Design for Partial Coordinated Transmission in Heterogenous Networks," in *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 226240, Feb. 2013.
- [11] Wan Ariffin, W.N.S.F., Zhang, X. and Nakhai, M.R., "Real-time Energy Trading with Grid in Green Cloud-RAN", in *IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, pp. 748-752, Sept. 2015.
- [12] P. Blasco and D. Gunduz, "Learning-Based Optimization of Cache Content in a Small Cell Base Station", in *IEEE International Conference on Communications*, June 2014.
- [13] W. Chen, Y. Wang, and Y. Yuan, "Combinatorial multi-armed bandit: General framework, results and applications," in *International Conference on Machine Learning*, Atlanta, GA, Jun. 2013.
- [14] T. A. Le, M. R. Nakhai, "A decentralized downlink beamforming algorithm for multicell processing", *IEEE Global Communications Conference*, 3-7 Dec. 2012.
- [15] M. Grant and S. Boyd, "CVX: Matlab Software for Disciplined Convex Programming, Version 2.0 (Beta)," <http://cvxr.com/cvx/>, Mar 2013.